Available online at www.sciencedirect.com
ScienceDirect
International Journal for Parasitology 37 (2007) 1297–1305
www.elsevier.com/locate/ijpara

Current Opinion

Tick genomics: The *Ixodes* genome project and beyond

J. Pagel Van Zee [a], N.S. Geraci [a], F.D. Guerrero [b], S.K. Wikel [c], J.J. Stuart [a], V.M. Nene [d], C.A. Hill [a,*]

[a] *Purdue University, 901 West State Street, West Lafayette, IN 47907, USA*
[b] *USDA-ARS Knipling-Bushland, Kerrville, TX 78028, USA*
[c] *University of Connecticut Health Center, Farmington, CT 06030, USA*
[d] *J. Craig Venter Institute, Rockville, MD 20850, USA*

Received 3 April 2007; received in revised form 18 May 2007; accepted 21 May 2007

Abstract

Ticks and mites (subphylum Chelicerata; subclass Acari) include important pests of animals and plants worldwide. The *Ixodes scapularis* (black-legged tick) genome sequencing project marks the beginning of the genomics era for the field of acarology. This project is the first to sequence the genome of a blood-feeding tick vector of human disease and a member of the subphylum Chelicerata. Genome projects for other species of Acari are forthcoming and their genome sequences will likely feature significantly in the future of tick research. Parasitologists interested in advancing the field of tick genomics research will be faced with specific challenges. The development of genetic tools and resources, and the size and repetitive nature of tick genomes are important considerations. Innovative approaches may be required to sequence, assemble, annotate and analyse tick genomes. Overcoming these challenges will enable scientists to investigate the genes and genome organisation of this important group of arthropods and may ultimately lead to new solutions for control of ticks and tick-borne diseases.
© 2007 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

Keywords: Acari; Ixodida; Ixodidae; *Ixodes scapularis*; Ticks; Genomics

1. Introduction

Ticks and mites (subphylum Chelicerata; subclass Acari) include parasites and pests of animals and plants worldwide and numerous species are of medical, veterinary or agricultural significance. Despite their global importance, our knowledge of the biology of the Acari is limited. This is due in part to the fact that many Acari are difficult to culture and are genetically intractable. Advances in acarology have also been stymied by lack of nucleotide sequence data. This is in contrast to the class Insecta for which there are numerous completed or ongoing genome sequencing projects, including at least eight projects for insect vectors of human diseases. Recently, large-scale sequencing projects have been initiated for two species of Acari, namely the black-legged tick *Ixodes scapularis* and the two-spotted spider mite *Tetranychus urticae*, and it is now becoming apparent that genome sequence data will play a significant and ever-increasing role in many areas of tick and mite research.

Unfortunately, a comprehensive review of genomics as it pertains to the Acari is beyond the scope of this paper. Our discussions are largely restricted to hematophagous, animal-parasitic species and more specifically to the ticks (superorder Parasitiformes; order Ixodida) in the families Ixodidae (hard ticks) and Argasidae (soft ticks). Ticks are of vast medical and veterinary importance; they transmit a greater variety of infectious agents than any other blood-feeding arthropod (Table 1) and they cause direct damage to their host through attachment and feeding (Jongejan and Uilenberg, 2004; Dennis and Piesman, 2005). Given their importance, it is surprising how little we know about many of the basic biological processes of ticks. For

* Corresponding author. Tel.: +1 765 496 6157; fax: +1 765 496 1219.
E-mail address: hillca@purdue.edu (C.A. Hill).

0020-7519/$30.00 © 2007 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijpara.2007.05.011

Table 1
Geographic distribution of ixodid and argasid ticks of medical and veterinary significance[a]

Geographic region	Species	Disease and causative agent
Africa	*Rhipicephalus appendiculatus*	East Coast Fever and Corridor Disease (*Theileria parva*); Benign Bovine Theileriosis (*Theileria taurotragi*); Nairobi Sheep Disease; Thogoto Disease
	Amblyomma variegatum	Cowdriosis (*Ehrlichia ruminantium*); African Tick-Bite Fever (*Rickettsia africae*); Benign African Theileriosis (*Theileria mutans*); Thogoto Disease
Africa/Eurasia	*Ixodes ricinus*	TBE; LD (*Borrelia burgdorferi* s.l.); Human (*Babesia microti*) and Bovine Babesiosis (*Babesia divergens*); HGA (*Anaplasma phagocytophilum*); Ruminant Tick-Borne Fever; *Rickettsia helvetica*; Louping-ill in sheep
	Hyalomma marginatum	Crimean-Congo Hemorrhagic Fever; Bovine Tropical Theileriosis (*Theileria annulata*)
	Ornithodoros moubata	African Human Relapsing Fever (*Borrelai duttonii*); African Swine Fever
Eurasia	*Ixodes persulcatus*	TBE; LD
	Dermacentor marginatus	Canine Babesiosis (*Babesia canis*); Human Rickettsiosis (*Ricketsia slovaka*), Q-Fever (*Coxiella burnetii*); Botonneuse Fever (*Rickettsia conorii*)
	Dermacentor reticulatus	Equine Babesiosis (*Babesia caballi*); Canine Babesiosis (*Babesia canis*); Siberian Tick Typhus (*Rickettsia sibirica*); Human Rickettsiosis (*Rickettia slovaka*); Botonneuse Fever
Europe	*Ixodes hexagonus*	LD
Asia	*Haemaphysalis spinigera*	Kyasanur Forest Disease
Eurasia, Australasia	*Haemaphysalis longicornus*	Bovine Babesiosis (*Babesia ovata*); East Asian Bovine Theileriosis (*Theileria buffeli*); Canine Babesiosis (*Babesia gibsoni*); Human Rickettsiosis (*Rickettsia japonica*)
Australia	*Ixodes holocyclus*	Tick paralysis in humans and animals (holocyclotoxin); *Rickettsia australis*
North America	*Ixodes scapularis*	LD; HGA & Equine Erlichiosis (*Anaplasma phagocytophilum*); Human Babesiosis; POW
	Ixodes pacificus	LD; HGA
	Dermacentor andersoni	POW; CTF; Bovine Anaplasmosis; RMSF (*Rickettsia rickettsii*); paralysis toxin
	Ornithodoros hermsi	TBRF (*Borrelia hermsii*)
North and Central America	*Dermacentor variabilis*	Bovine Anaplasmosis; RMSF; paralysis toxin in animals and humans; Tularemia
North, Central and South America	*Amblyomma americanum*	Tularemia; HME (*Ehrlichia chaffeensis*); HGE (*Ehrlichia ewingii*); STARI (*Borrelia lonestari*);
	Amblyomma cajennense	RMSF
	Amblyomma maculatum	Experimental vector of Cowdriosis; Hepatozoonosis (*Hepatozoon americanum*)
Africa, Americas, Eurasia	*Rhipicephalus (Boophilus) annulatus*	Bovine Babesiosis (*Babesia bovis, Babesia bigemina*); Bovine Anaplasmosis
	Rhipicephalus sanguineus	Canine Ehrlichiosis (*Ehrlichia canis*); Canine Babesiosis (*Babesia vogeli*); Canine Hepatozoonosis (*Hepatozoon canis*); Human Tick-Bite Fever (*Rickettsia conorii*); HGE
Africa, Americas, Australia	*Rhipicephalus (Boophilus) microplus*	Bovine Babesiosis (*B. bovis, B. bigemina*); Bovine Anaplasmosis; Equine Piroplasmosis (*Theileria equi*)

[a] After Jongejan and Uilenberg (2004); CTF, Colorado Tick Fever; HGA, Human Granulocytic Anaplasmosis; HGE, Human Granulocytic Ehrlichiosis; HME, Human Monocytic Ehrlichiosis; LD, Lyme Disease; POW, Powassan Encephalitis; RMSF, Rocky Mountain Spotted Fever; STARI, Southern Tick-associated Rash Illness; TBE, Tick-borne Encephalitis; TBRF, Tick-borne Relapsing Fever.

example, the molecular basis of tick sensory perception, tick-host-pathogen interactions, blood meal digestion, oogenesis, reproduction and development are not well known. Furthermore, the genetic basis of important phenotypes such as tick-host preference, vector competence for specific pathogens and acaricide resistance are poorly understood. Genome sequencing has been embraced as a method to address the above shortcomings. Parasitologists must be ready to exploit large amounts of sequence data from the *Ixodes* project and other tick sequencing projects that will ultimately follow. However, genome studies of the Acari and the Ixodida, in particular, will not be without challenges. Our goal in this publication is to position the field of tick and mite research to take advantage of genomic data. We review the current status of tick genomics, discuss realistic expectations for tick genome projects and highlight areas for investment of future research efforts.

2. Genomics of the Acari: an emerging field

Several recent developments signal that the field of acarology has entered the genomics era (Table 2). One of the most pivotal is the *I. scapularis* genome project (IGP). *Ixodes scapularis*, the black-legged tick or deer tick, is a hard (ixodid) tick vector of the causative agents of Lyme disease, babesiosis and anaplasmosis in the United States. The IGP was initiated in 2004, its goal being to sequence the genome of a medically significant tick (Hill and Wikel, 2005). The project represents two important scientific firsts; it is the first sequencing project for a tick and a chelicerate. The

Table 2
Ongoing and proposed genome projects for the subclass Acari

Species	Genome size	In-bred strain[a]	Sex determination[b]	No. of ESTs	BAC library[c]	WGS[d]
Ongoing projects						
Ticks (Parasitiformes)						
Ixodes scapularis (black-legged tick)	2.1 Gbp[e]	Yes	XX, XY	240,000[f]	10X	6X (Wikel strain)
Mites (Acariformes)						
Tetranychus urticae (two-spotted spider mite)	75 Mbp[g]	Yes	Haplo-diploid	75,000[h]		8X (London strain)
						1X (France strain)
Proposed projects						
Hard ticks (Ixodidae)						
Rhipicephalus microplus (Southern cattle tick)	7.1 Gbp[e]	Yes	XX, XO	~50,000	1X	–
Dermacentor variabilis (American dog tick)	2.91 Gbp[i]	Yes	XX, XO	–	–	–
Amblyomma americanum (lone star tick)	3.18 Gbp[i]	Yes	XX, XO	–	–	–
Soft ticks (Argasidae)						
Ornithodoros moubata/hermsi	ND	Yes	XX, XY	–	–	–
Mites						
Leptotrombidium spp.	ND	Yes	–	–	–	–

Abbreviations: BAC, bacterial artificial chromosome; EST, expressed sequence tag; Gbp, giga basepairs; Mbp, mega basepairs; ND, not determined; WGS, whole genome shotgun sequencing.
[a] The exact number of in-bred generations has not been determined for these colonies.
[b] After Oliver (1977).
[c] Estimated BAC clone coverage of genome.
[d] Proposed WGS genome coverage.
[e] After Ullmann et al. (2005).
[f] Approved as part of NIH funded project.
[g] After Dearden et al. (2002).
[h] Approved as part of DOE funded project.
[i] After Geraci et al. (2007).

project is funded by the National Institutes of Health (NIH), USA and is a collaboration between the international tick research community and two genome sequencing centres, namely the J. Craig Venter Institute (JCVI), USA and the Broad Institute, USA. Like all genome projects, the IGP is a multi-phase and multi-investigator undertaking. Current plans call for whole genome shotgun sequencing (WGS) to approximately sixfold coverage of the genome. Trace reads from small (2–4 Kbp), medium (10–12 Kbp) and large (40–50 Kbp) insert genomic clones will be the basis for assembly of the genome sequence. Also included are reads of ~160,000 bacterial artificial chromosome (BAC) clone ends (BAC-end sequencing), the complete end-to-end sequencing of 60 BAC clones and ~240,000 expressed sequence tag (EST) reads. Paired BAC end reads span large segments of the genome and will be used to help assemble WGS sequence into scaffolds. Sequenced BACs will provide an early insight into the *Ixodes* genome and will have utility as probes for physical mapping of assembled sequence to chromosomes. ESTs are arguably one of the most valuable resources generated as part of any genome project. They will be used to identify expressed genes, confirm gene predictions and to train automated gene finding algorithms.

More than 18 million *Ixodes* trace reads have been deposited at the National Center for Biotechnology Information (NCBI) trace archive to date; this represents more than fivefold coverage of the genome. Sequencing is ongoing and assembly and annotation based on at least sixfold coverage is expected by sometime in 2008. Preliminary analysis of an assembly of WGS data has revealed important information about the feasibility of the genome project; all indications are that a useful assembly will be produced at higher coverage. To date, 20 *Ixodes* BAC clones have been shotgun sequenced and assembled. Also available are approximately 370,000 BAC-end reads and more than 80,000 ESTs have been sequenced from a normalised pooled stage/tissue library. All sequences have been deposited at NCBI. Tentative consensus (TC) sequences have been produced from the ESTs and both TCs and singletons can be accessed through the *I. scapularis* Gene Index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=i_scapularis). The *Ixodes* mitochondrial (mt) genome has been assembled and annotated; the genome resembles published mt genomes from other *Ixodes* species (Shao et al., 2005). The WGS data also contains sequences which cluster with *Rickettsia* spp. sequences. Assembling these sequences will enable scien-

tists to investigate the role of this probable symbiont in *Ixodes* biology and its potential for use in paratransgenesis.

Another exciting recent development is the approval of the *T. urticae* genome project by the Department of Energy (http://www.jgi.doe.gov/sequencing/why/CSP2007/spidermite.html). This project will shotgun sequence the *T. urticae* "London" strain to eightfold coverage and the "France" strain to onefold coverage. *T. urticae* (superorder Acariformes) is a significant pest of agricultural crops worldwide and the project will generate much needed nucleotide data for a phytophagous mite species. The project is also an important resource for tick researchers because parallel analyses between the superorders Acariformes and Parasitiformes may reveal pathways of genes essential for plant versus animal parasitism.

3. Beyond the *Ixodes* and *Tetranychus* genome projects

An emerging trend in the field of genomics is that sequencing projects are no longer limited to a single representative species but now often include multiple organisms within a phylogenetic group. Thus, it is not unreasonable to assume that genome projects for other tick and mite species will follow on the heels of the *Ixodes* and *Tetranychus* projects.

USDA-ARS scientists have generated approximately 50,000 ESTs and a BAC library for the hard tick vector of bovine babesiosis, *Rhipicephalus* (*Boophilus*) *microplus* in anticipation of a *R. microplus* genome project (Guerrero et al., 2006). *Rhipicephalus microplus* is a member of the metastriate lineage of ticks which includes numerous genera and species of medical and veterinary importance. In comparison, *I. scapularis* is a member of the prostriate lineage which comprises the single genus *Ixodes*. The pro- and metastriate lineages differ markedly in many aspects of their biology such as type of developmental cycle (i.e., three- versus one-host ticks), host range and vector competence. Comparative analyses between pro- and metastriate genomes may reveal the genetic basis for fundamental differences in the biology of these lineages.

At a meeting organised by the NIH in November 2006, additional tick and mite species were considered for future genome projects (Table 2). Proposed species include two additional metastriate ticks, namely *Demacentor variabilis* (American dog tick) and *Amblyomma americanum* (lone star tick) that vector *Rickettsia* and *Ehrlichia* spp. bacteria, respectively. Also considered was a soft tick (family Argasidae; possibly *Ornithodoros moubata* or *Ornithodoros hermsi*) vector of tick-borne relapsing fever, and a *Leptotrombidium* spp. of mite that vectors the causative agent of scrub typhus. It remains to be seen which, if any of these species will be approved, and certainly there may be other species that prove more attractive for sequencing. It should be noted that key resources such as BAC, genomic and cDNA libraries are not in place for many of these "proposed" species (Table 2). Production of genome project resources is largely the responsibility of the research communities interested in sequencing a particular organism. Development of such resources can be time consuming and expensive but it is an invaluable investment and should commence forthwith if these proposed projects are ever to become a reality.

4. Specific challenges associated with tick genomics

As the flagship tick genome project, the IGP will be an important model for subsequent tick and mite sequencing efforts. The picture that is now emerging is that the IGP and subsequent tick genome projects will present sequencing centres and parasitologists interested in sequence data with specific challenges. There are several reasons for this. Firstly, the ticks (at least those of interest for genome studies) appear to have large genomes composed of significant amounts of repetitive DNA which is potentially problematic for genome assembly. Second, WGS based on the Sanger sequencing method (Sanger et al., 1977) relies on inbred colonies with minimal polymorphism, which for the most part are not available for many Acari. Third, considerable evolutionary distance between ticks and insects may limit homology-based gene predictions. Lastly, the genomic tools and resources that go hand-in-hand with WGS projects are not in place for many ticks and mites. While challenging, these issues are by no means unique, nor in many cases will they be insurmountable. Nonetheless, they warrant a detailed examination not only to provide scientists with realistic expectations of tick genomics but also to encourage appropriate development of this field.

4.1. Large genomes

Preliminary knowledge of genome size is desirable before commitment to WGS because genome size determines the amount of sequencing required, and hence the cost of a sequencing project. The genome size of *T. urticae* has been estimated at 75 Mbp (Dearden et al., 2002) but extensive studies of mite genome size have not been conducted to date. Several studies suggest that the Ixodida have large haploid genomes. Using re-association kinetics, the genome sizes of *I. scapularis* and *R. microplus* were estimated to be approximately 2.1 and 7.1 Gbp, respectively, (Ullmann et al., 2005) and the *A. americanum* genome was estimated to be approximately 1.04 Gbp (Palmer et al., 1994). Geraci et al. used flow cytometry to evaluate genome size in a range of argasid and ixodid ticks. The average haploid genome size of the Argasidae was estimated to be 1.28 Gbp and that of the Ixodidae was 2.67 Gbp. The implications of these findings are that extensive sequencing of the Ixodida listed in Table 2 will be necessary in order to achieve adequate genome coverage. Shotgun sequencing of these species at today's costs will be an expensive endeavor and WGS may only be feasible to a draft level of coverage. Fortunately, sequencing costs continue to decline. New sequencing technologies such as the 454 (Roche, Basel, Switzerland; Margulies et al.,

2005) and Solexa (Illumina, Inc., San Diego, CA) sequencing platforms (Bentley, 2006) may eventually offer a more efficient and cost effective solution for generation of draft sequence. The utility of 454 sequencing for large, complex eukaryote genomes is being considered (Wicker et al., 2006). At the very least, these technologies may be used to analyse genetic variation in ticks or to improve draft sequence generated by the traditional Sanger method (Goldberg et al., 2006). Strategies such as methylation filtration and re-association kinetics (high C_oT selection) which enrich the proportion of unique, presumably coding sequence in a DNA sample, may also help to get at the "gene space" in species with large genomes (Meyers et al., 2004). These have proved to be valid alternatives to the traditional Sanger sequencing approach in maize (Palmer et al., 2003; Whitelaw et al., 2003). For some tick species, deep sequencing of cDNA libraries may offer an attractive alternative to WGS although these approaches will not provide an assembled genomic sequence.

At sixfold genome coverage, scientists investigating the *I. scapularis* sequence will have to contend, at least initially, with a lower level of sequence resolution for *Ixodes* than for other arthropods sequenced to date. This will translate into a greater number of assembled supercontigs (or scaffolds) of shorter length which may be problematic for genome annotation. As the first and only tick genome project currently approved for sequencing, *Ixodes* is an important model organism for acarine genomics and tick research. At ~2.1 Gbp, *Ixodes* has a haploid genome approximately two-thirds the size of the human genome but it has a smaller genome than other ixodid ticks of human and animal health significance listed in Table 2. Additional sequencing of *Ixodes* to a higher level of genome coverage would be a worthy and warranted investment.

4.2. Repetitive DNA

Large eukaryotic genomes are known to comprise significant amounts of non-coding, repetitive DNA. Given the results of Geraci et al. (2007), it is a fair assumption that this will hold true for many hard and soft ticks. An understanding of the classes, amounts and distribution of repetitive DNA in a genome is an important outcome of genome sequencing. Highly repetitive DNA of low sequence complexity such as tandem and dispersed repeats can complicate genome assembly although it can often be successfully masked to enhance the assembly. C_oT studies have already shown that a significant percentage of the *I. scapularis*, *R. microplus* and *A. americanum* genomes are comprised of highly repetitive sequences (Palmer et al., 1994; Ullmann et al., 2005).

Genomic sequence generated independently of the IGP has been used to identify highly and moderately repetitive DNA in *I. scapularis* and *R. microplus* (Fig. 1; Hill, C.A., personal communication) using the repeat finder program RECON 1.05 (Bao and Eddy, 2002). Repetitive DNA comprised a significant proportion of the *Rhipicephalus* and *Ixodes* sequences (approximately 15.7% and 21.5%, respectively). Most repeat elements appear to be present in relatively low copy numbers in the *Rhipicephalus* and *Ixodes* genomes and are probably not of concern for assembly. However, some repeats exist in very high copy numbers, especially in *Ixodes* and likely constitute a significant proportion of the repetitive fraction. One example of note is a ~95 bp tandem repeat tentatively named "Is6" that we have identified from *I. scapularis* genomic DNA (gDNA). We estimate that the *Ixodes* genome contains more than 1 million copies of Is6, which appears to be highly conserved and is predicted to constitute approximately 5% of the genome sequence. This supports earlier reports of low sequence complexity in ticks and highlights important species-specific differences in repetitive DNA content.

More recently, we have used fluorescent in situ hybridisation (FISH) to physically map several classes of abundant tandem and dispersed repeats to *Rhipicephalus* chromosomes (Hill et al., unpublished data). Many highly abundant repeats localise to the telomeres of chromosomes. These data indicate that assembly of large, repetitive tick genomes will not be a trivial undertaking. One problem encountered by many eukaryote genome projects is that some genome regions such as the telomeres often do not assemble well, even with repeat masking. Regardless, unassembled trace sequence is a major scientific resource for gene discovery and population genetics studies and even a fragmented assembly will have application for analysis of genome architecture. Fortunately, the *Tetranychus* project may prove helpful for assembly and annotation of large, repeat-rich tick genomes. At ~75 Mbp, *T. urticae* has one of the smallest invertebrate genomes identified to date (Dearden et al., 2002) and the genome is expected to contain limited repetitive sequence. The *T. urticae* assembly is expected to comprise large scaffolds with minimal gaps and fewer unassembled regions of the genome. As such, it may provide an important framework for assembly of a more fragmented *Ixodes* genome composed of numerous, shorter scaffolds. Furthermore, it may generate a near complete set of gene model predictions for *Tetranychus* with application for automated gene builds in other Chelicerates. Ultimately, *Ixodes–Tetranychus* whole genome comparisons will advance understanding of genome organisation and evolution across the subclass Acari.

4.3. The need for in-bred lines

The traditional WGS approach to genome sequencing and assembly relies on genomic sequence with low levels of polymorphism. Multiple factors are considered when selecting a strain for a genome sequencing effort (e.g., colony robustness and ease of culture, relevance of phenotypes and availability of tools and resources to support accompanying genome studies) but one of the most important is whether a strain has been sufficiently in-bred to minimise polymorphism. Microgram quantities of purified,

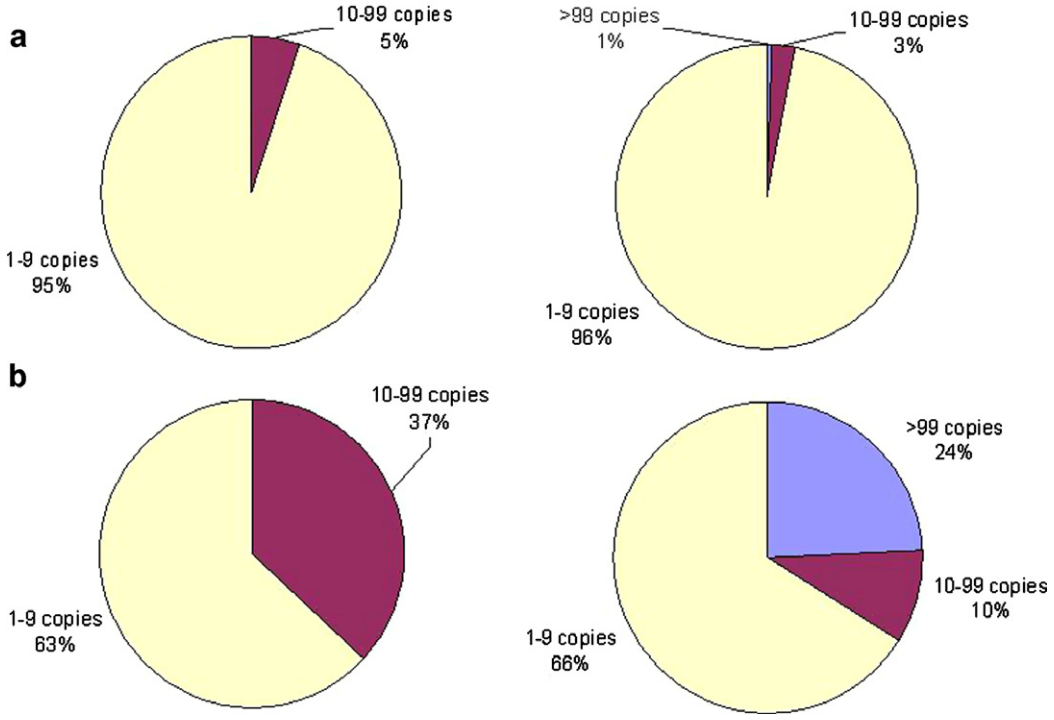

Fig. 1. *Rhipicephalus microplus* (left) and *Ixodes scapularis* (right) repetitive DNA. The repeat finder program RECON 1.05 was used to analyse 662,404 bp of *R. microplus* and 638,000 bp of *I. scapularis* genomic sequences. The number of copies (one to nine, 10–99 or greater than 99 copies) of a repetitive element per repeat family is shown as a factor of: (a) the total number of repeat families identified per species and (b) the total number of nucleotide bases analysed per species.

high molecular weight gDNA are generally required for production of libraries for sequencing. Because invertebrates typically yield only modest amounts of gDNA, most invertebrate genome projects use gDNA pooled from multiple individuals of an in-bred laboratory strain. Haplotype issues are generally overcome by sequencing to a relatively high level of coverage such that a dominant haplotype emerges. Unfortunately, the long life-cycle of some tick species, as well as the need to supplement with field collected material in order to maintain colony fitness, has been a major impediment to the development of truly in-bred tick lines (i.e., lines established from a single-pair mating). The *I. scapularis* Wikel strain selected for the IGP was established from field material and has been in-bred in the laboratory for more than 12 generations. This is a significant feat considering that the life-cycle of *I. scapularis* is completed over 2 years in the field and can only be shortened, at best, to 9 months under optimal conditions in the laboratory. Scientists interested in sequencing a particular tick or mite species are encouraged to plan ahead by establishing in-bred lines as soon as possible because these are likely to factor significantly into funding decisions. Both arrhenotoky and thelytoky have been noted in the Acari (Sonenshine, 1991); these forms of parthenogenesis may offer a convenient biological mechanism to reduce haplotype variation in a sequenced population. That said, for some high priority species, in-bred lines and/or reduced levels of polymorphism may never be a reality. In such cases, revolutionary changes in sequencing methods and assembly algorithms may be the only realistic options for generation of useful assemblies.

4.4. Sequence divergence and unique gene architecture

The subphylum Chelicerata which comprises the spiders, scorpions, horseshoe crabs, ticks and mites is expected to have arisen during the Cambrian Epoch (Brusca and Brusca, 1990). Based on the fossil record (Benton, 1993), it may have been some 490–550 million years (Myr) since the lineages which gave rise to ticks and insects shared a common ancestor (Douzery et al., 2004). Hoogstraal suggested that ancestral ticks, resembling the present-day Argasidae, arose in the late Paleozoic or early Mesozoic approximately 225 Myr ago on reptiles, while Oliver suggested that ticks may have originated as early as the Devonian, approximately 350–400 Myr ago on amphibians (Klompen et al., 1996). Lack of homology between tick and insect genes will thus present a challenge for homology-based tick gene predictions, especially with respect to highly divergent gene families. Amino acid sequence similarity between mosquito and tick opsin genes highlights this point. The opsins are a family of putative visual receptors that are highly conserved between mammals and invertebrates. Amino acid sequence similarity between opsin orthologs of the malaria mosquito *Anopheles gambiae* and the yellow fever mosquito *Aedes aegypti*,

two dipterans expected to have shared a common ancestor approximately 145–200 Myr ago (Krzywinski et al., 2006), ranges from 50% to 89%. The highest amino acid sequence similarity that has been identified to date between what is likely the major opsin in *Ixodes* and the closest ortholog in *Aedes* is only 52%. Low gene density and a range of intron sizes have been noted in assembled *Ixodes* BAC sequence. These features may ultimately translate into inaccurate and/or split gene models and, therefore, an over-prediction of the proteome.

Unless they are associated with an EST, it is highly probable that many divergent tick genes will not be identified from assembled sequences or that they will be inaccurately annotated. Tick EST sequences will be an essential investment for scientists contemplating a genome project. Community expertise in manual tick genome annotation will also be invaluable and will be facilitated by the number and diversity of sequenced and annotated insect genomes that are now available through NCBI and dedicated databases such as VectorBase (Lawson et al., 2007; http://www.vectorbase.org/index.php). Unfortunately, the EST approach cannot be used to identify the full compliment of genes, transcripts and proteins – this goal is one that can only be realised with an assembled genome.

4.5. Genetic and physical mapping

Typically, large, repetitive genomes cannot be assembled based on shotgun sequence alone. Genetic (linkage) and physical maps are an important component of most genome sequencing efforts because genetic markers and physical mapping can assist assembly of the genomic sequence and the assignment and orientation of sequence to chromosomes. The goal of most genome projects initiated since the human genome project is the integration of genetic and physical mapping data with genome sequence data (Meyers et al., 2004). Linkage maps also complement genomic sequence as they can be used to identify regions of the genome associated with phenotypes of interest (i.e., QTL mapping). While various types of molecular markers have been developed for a range of ticks and mites (reviewed by Navajas and Fenton, 2000), only one linkage map has been published for an acarine species to date. This is the preliminary map of *I. scapularis* produced by Ullmann et al. (2003) and based on 84 random amplified polymorphic (RAPD) markers, 32 sequence-tagged RAPD (STAR) markers, five cDNAs and five microsatellites. This work demonstrates the feasibility of linkage mapping for ticks but second generation, high resolution linkage maps are required for *Ixodes* and other pro- and metastriate ticks.

A physical map is a linearly ordered set of DNA fragments separated by a known physical distance (expressed in bp). Physical maps can be particularly useful when contiguous segments of assembled sequence do not overlap due to low genome coverage or repetitive sequence. Most physical maps are currently based on large-insert clones such as BACs. The technique of BAC fingerprinting which employs genome-wide restriction digest patterns to order overlapping BACs into contigs can be used to identify clones that span predicted gaps (Meyers et al., 2004). Although BAC fingerprinting and contiging are affected by genome size and highly repetitive DNA, they may have applications for sequence assembly in *Ixodes* and other tick species. Techniques such as optical mapping may also be on the horizon for mapping of large, complex eukaryote genomes (Valouev et al., 2006). An assembled sequence must also be ordered and placed on chromosomes. Recently, we have developed a technique for FISH mapping of labeled BAC DNA to meiotic chromosome spreads of *R. microplus* (Fig. 2). This approach is being used to study the distribution and abundance of repetitive DNA in the *R. microplus* genome and to identify specific chromosomes. These techniques will have applications in tick genome assembly and population genetics studies; it is imperative that they be extended to other species of ticks and mites.

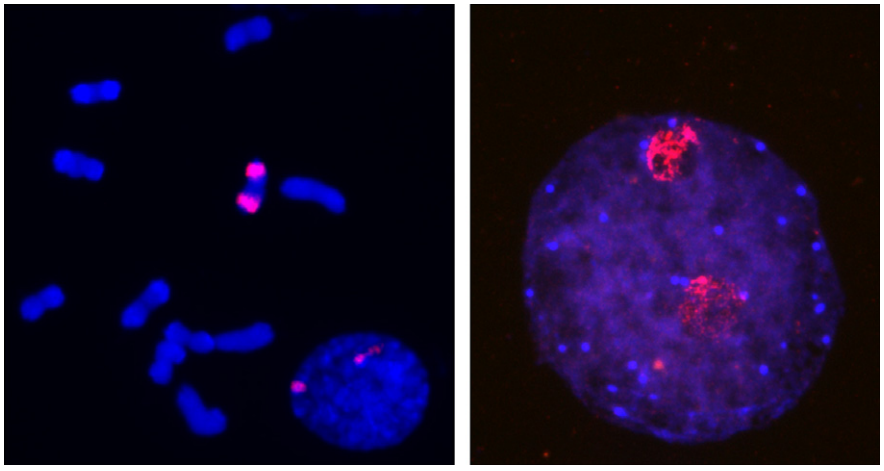

Fig. 2. Fluorescent in situ hybridisation (FISH) mapping of a ribosomal DNA (rDNA) probe (red label) to *Rhipicephalus microplus* meiotic chromosomes (left panel) and diploid interphase nucleus (right panel) (Ullmann et al., in press). With kind permission of Springer Science and Business Media.

5. Conclusion: tick genomics – the pros outweigh the cons

The challenges associated with the sequencing and analysis of tick genomes are far outweighed by the opportunities afforded by an assembled sequence. Data from the *Ixodes* project provides scientists with the first opportunity to investigate the nature and genome organisation of this important group of arthropods. As with all genome projects, gene identification will be an obvious focus for many research groups. Genes can reveal the molecular basis of important biological processes and can even expose new pathways that would not otherwise have been discovered using an EST approach. Gene identification will also be driven by the pressing need to find new targets for vaccine and acaricide development. Perhaps less obvious, but no less intriguing, will be the opportunity to study the architecture of tick genomes and gene/genome evolution in the Acari. Chelicerates occupy a basal phylogenetic position within the phylum Arthropoda and are thus central to understanding the evolution of arthropod lineages (Grbic et al., 2007). An assembled sequence will facilitate population genetic studies in *Ixodes* and other tick species. It will also provide a much needed platform for proteomic analyses of the Ixodida.

As is the case with all genome projects, it is certain that we will be refining and exploiting the *Ixodes* genome sequence for many years to come. Given the size of most tick genomes, shotgun sequencing of additional tick species will be costly. However, with reductions in sequencing costs and the introduction of new genome sequencing methods, it is almost certain that such projects will become a reality. In the interim, researchers are encouraged to invest in developing laboratory strains of ticks and mites that are suitable for shotgun sequencing and the genomic resources and expertise needed to accompany a genome sequencing effort.

Acknowledgement

We thank B. Birren of the Broad Institute of MIT/Harvard for contributions to the *Ixodes* sequencing effort and for helpful comments on this manuscript.

References

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–1276.

Bentley, D.R., 2006. Whole genome re-sequencing. Curr. Opin. Genet. Dev. 6, 545–552.

Benton, M.J., 1993. The Fossil Record 2. Chapman and Hall, London.

Brusca, R.C., Brusca, G.J., 1990. Invertebrates. Sinauer Associates, Inc, Sunderland, MA.

Dearden, P.K., Donly, C., Grbic, M., 2002. Expression of pair-rule homologues in a chelicerate: early patterning of the two-spotted spider mite *Tetranychus urticae*. Development 129, 5461–5472.

Dennis, D.T., Piesman, J.F., 2005. Overview of tick-borne infections of humans. In: Goodman, J.L., Dennis, D.T., Sonenshine, D.E. (Eds.), Tick-borne Diseases of Humans. American Society for Microbiology, Washington, pp. 3–11.

Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA 101, 15386–15391.

Geraci, N.S., Johnston, J.S., Robinson, J.P., Wikel, S.K., Hill, C.A., 2007. Variation in genome size of ixodid and argasid ticks. Insect Biochem. Mol. Biol. 37, 399–408.

Goldberg, S.M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S.A., Lauro, F.M., Li, K., Rogers, Y.H., Strausberg, R., Sutton, G., Tallon, L., Thomas, T., Venter, E., Frazier, M., Venter, J.C., 2006. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc. Natl. Acad. Sci. USA 103, 11240–11245.

Grbic, M., Khila, A., Lee, K-Z., Bjelica, A., Grbic, V., Whistlecraft, J., Verdon, L., Navajas, M., Nagy, L., 2007. Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism. BioEssays 29, 1–8.

Guerrero, F.D., Nene, V.M., George, J.E., Barker, S.C., Willadsen, P., 2006. Sequencing a new target genome: the *Boophilus microplus* (Acari: Ixodidae) genome project. J. Med. Ent. 43, 9–16.

Hill, C.A., Wikel, S.K., 2005. The *Ixodes scapularis* genome project: an opportunity for advancing tick research. Trends Parasitol. 21, 151–153.

Jongejan, F., Uilenberg, G., 2004. The global importance of ticks. Parasitol. 129, S3–S14.

Klompen, J.S., Black IV, W.C., Keirans, J.E., Oliver Jr., J.H., 1996. Evolution of ticks. Ann. Rev. Entomol. 41, 141–161.

Krzywinski, J., Grushko, O.G., Besansky, N.J., 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol. Phylogenet. Evol. 39, 417–423.

Lawson, D., Atkinson, P., Hammond, M., Bruggner, R., MacCallum, R., Redmond, S., Megy, K., Besansky, N., Dialynas, E., Hill, C., Campbell, K., Emmert, D., Russo, S., Christophides, G., Birney, E., Gelbart, W., Louis, C., Kafatos, F., Severson, D., Collins, F., 2007. Vectorbase: a home for invertebrate vectors of human pathogens. Nucleic Acids Res. 35, D503–D505.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, XV., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Meyers, B.C., Scalabrin, S., Morgante, M., 2004. Mapping and sequencing complex genomes: let's get physical! Nat. Rev. Genet. 5, 578–588.

Navajas, N., Fenton, B., 2000. The application of molecular markers in the study of diversity in acarology: a review. Expt. Appl. Acarol. 24, 751–774.

Oliver, J.H., 1977. Cytogenetics of mites and ticks. Ann. Rev. Entomol. 22, 407–429.

Palmer, M.J., Bantle, J.A., Guo, X., Fargo, W.S., 1994. Genome size and organization in the ixodid tick *Amblyomma americanum*. Insect Mol. Biol. 3, 57–62.

Palmer, L.E., Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V.S., Nascimento, L.U., Dike, S., de la Bastide, M., Martienssen, R.A., McCombie, W.R., 2003. Maize genome sequencing by methylation filtration. Science 302, 2115–2117.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Shao, R., Barker, S.C., Mitani, H., Aoki, Y., Fukunaga, M., 2005. Evolution of duplicate control regions in the mitochondrial genomes of

Metazoa: a case study with Australasian *Ixodes* ticks. Mol. Biol. Evol. 22, 620–629.

Sonenshine, D.E., 1991. Biology of Ticks, Vol. 1. Oxford University Press, New York.

Ullmann, A.J., Stuart, J.J., Hill, C.A., in press. Tick. In: Cole, C.R. and Hunter, W. (Eds.), Genome Mapping in Animals, Vol. 3: Insects. Springer-Verlag, Heidelberg, Germany.

Ullmann, A.J., Piesman, J., Dolan, M.C., Black, IV W.C., 2003. A preliminary linkage map of the hard tick, *Ixodes scapularis*. Insect Mol. Biol. 12, 201–210.

Ullmann, A.J., Lima, C.M., Guerrero, F.D., Piesman, J., Black IV, W.C., 2005. Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. Insect Mol. Biol. 14, 217–222.

Valouev, A., Schwartz, D.C., Zhou, S., Waterman, M.S., 2006. An algorithm for assembly of ordered restriction maps from single DNA molecules. Proc. Natl. Acad. Sci. USA 103, 15770–15775.

Whitelaw, C.A., Barbazuk, W.B., Pertea, G., Chan, A.P., Cheung, F., Lee, Y., Zheng, L., van Heeringen, S., Karamycheva, S., Bennetzen, J.L., SanMiguel, P., Lakey, N., Bedell, J., Yuan, Y., Budiman, M.A., Resnick, A., Van Aken, S., Utterback, T., Riedmuller, S., Williams, M., Feldblyum, T., Schubert, K., Beachy, R., Fraser, C.M., Quackenbush, J., 2003. Enrichment of gene-coding sequences in maize by genome filtration. Science 302, 2118–2120.

Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B., Stein, N., 2006. 454 sequencing put to the test using the complex genome of barley. BMC Genomics 7, 275–285.